

37  
-3-

1380X

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: LABIGNE, AGNES  
CUSSAC, VALERIE  
FERRERO, RICHARD

(ii) TITLE OF INVENTION: GENES OF HELICOBACTER PYLORI NECESSARY  
FOR THE REGULATION AND MATURATION OF UREASE AND THEIR USE

(iii) NUMBER OF SEQUENCES: 12

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,  
P.C.

(B) STREET: 1755 S. Jefferson Davis Highway, Suite 400

(C) CITY: Arlington

(D) STATE: Virginia

(E) COUNTRY: U.S.A.

(F) ZIP: 22202

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk

(B) COMPUTER: IBM PC compatible

(C) OPERATING SYSTEM: PC-DOS/MS-DOS

(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: US 08/211,312

(B) FILING DATE: 01-JUL-1994

(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: FR 91 12198

(B) FILING DATE: 03-OCT-1991

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: PCT/FR92/00921

(B) FILING DATE: 02-OCT-1992

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Oblon, Norman F.

(B) REGISTRATION NUMBER: 24,618

(C) REFERENCE/DOCKET NUMBER: 660-075-0XPCT

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (703) 413-3000

(B) TELEFAX: (703) 413-2220

(C) TELEX: 248855 OPAT UR

38

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3559 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 2..16

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 211..795

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 800..1309

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1324..2091

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 2122..2718

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 2721..3515

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

A CTC TTT AGC ATT TTC TAGGATTTTT TAGGAGCAAC GCTCTTAGAT CCTTAGTTTT	56
Leu Phe Ser Ile Phe	
1 5	
TAGCTCTCTG ATTTTTTGTT TATCAAAAAA TTGGGGGCTT TTTTGTGTTT TATTTTTTGT	116
CAATTTACTA TTTTCTTTA TGATTAGCTC AAGCAACAAA AGTTATTCGT AAGGTGCGTT	176
TGTTGTAAAA ATTTTGTGTT GGAAGGAAAA GGCA ATG CTA GGA CTT GTA TTG	228
Met Leu Gly Leu Val Leu	
1 5	
TTA TAT GTT GGG ATT GTT TTA ATC AGC AAT GGG ATT TGC GGG TTA ACC	276
Leu Tyr Val Gly Ile Val Leu Ile Ser Asn Gly Ile Cys Gly Leu Thr	

29  
-8-

	10	15	20	
	AAA GTC GAT CCT AAA AGC ACT GCG GTG ATG AAC TTT TTT GTG GGT GGG			324
	Lys Val Asp Pro Lys Ser Thr Ala Val Met Asn Phe Phe Val Gly Gly			
	25	30	35	
	CTC TCC ATT ATT TGT AAT GTG GTT GTC ATC ACT TAT TCC GCT CTC AAC			372
	Leu Ser Ile Ile Cys Asn Val Val Ile Thr Tyr Ser Ala Leu Asn			
	40	45	50	
	CCT ACA GCC CCT GTA GAA GGT GCT GAA GAT ATT GCT CAA GTA TCA CAC			420
	Pro Thr Ala Pro Val Glu Gly Ala Glu Asp Ile Ala Gln Val Ser His			
	55	60	65	70
	CAT TTG ACT AAT TTC TAT GGG CCA GCG ACT GGG TTA TTG TTT GGT TTC			468
	His Leu Thr Asn Phe Tyr Gly Pro Ala Thr Gly Leu Leu Phe Gly Phe			
	75	80	85	
	ACC TAC TTG TAT GCG GCT ATC AAC CAC ACT TTT GGT TTG GAT TGG AGG			516
	Thr Tyr Leu Tyr Ala Ala Ile Asn His Thr Phe Gly Leu Asp Trp Arg			
	90	95	100	
	CCC TAC TCT TGG TAT AGC TTA TTC GTA GCG ATC AAC ACG ATT CCT GCT			564
	Pro Tyr Ser Trp Tyr Ser Leu Phe Val Ala Ile Asn Thr Ile Pro Ala			
	105	110	115	
	GCG ATT TTA TCC CAC TAT AGC GAT ATG CTT GAT GAC CAC AAA GTG TTA			612
	Ala Ile Leu Ser His Tyr Ser Asp Met Leu Asp Asp His Lys Val Leu			
	120	125	130	
	GGC ATC ACT GAA GGC GAT TGG TGG GCG ATC ATT TGG TTG GCT TGG GGT			660
	Gly Ile Thr Glu Gly Asp Trp Trp Ala Ile Ile Trp Leu Ala Trp Gly			
	135	140	145	150
	GTT TTG TGG CTT ACC GCT TTC ATT GAA AAC ATC TTG AAA ATC CCT TTA			708
	Val Leu Trp Leu Thr Ala Phe Ile Glu Asn Ile Leu Lys Ile Pro Leu			
	155	160	165	
	GGG AAA TTC ACT CCA TGG CTT GCT ATC ATT GAG GGC ATT TTA ACC GCT			756
	Gly Lys Phe Thr Pro Trp Leu Ala Ile Ile Glu Gly Ile Leu Thr Ala			
	170	175	180	
	TGG ATC CCT GCT TGG TTA CTC TTT ATC CAA CAC TGG GTG TGAG ATG ATC			805
	Trp Ile Pro Ala Trp Leu Leu Phe Ile Gln His Trp Val Met Ile			
	185	190	195	1
	ATA GAG CGT TTA ATA GGC AAT CTA AGG GAT TTA AAC CCC TTG GAT TTC			853
	Ile Glu Arg Leu Ile Gly Asn Leu Arg Asp Leu Asn Pro Leu Asp Phe			
	5	10	15	
	AGC GTG GAT TAT GTG GAT TTG GAA TGG TTT GAA ACG AGG AAA AAA ATC			901
	Ser Val Asp Tyr Val Asp Leu Glu Trp Phe Glu Thr Arg Lys Lys Ile			

40

20	25	30	
GCT CGC TTT AAA ACC AGG CAA GGC AAA GAC ATA GCC GTA CGC CTT AAA Ala Arg Phe Lys Thr Arg Gln Gly Lys Asp Ile Ala Val Arg Leu Lys 35 40 45 50			949
GAC GCT CCC AAG TTG GGT TTC TCT CAA GGA GAT ATT TTA TTT AAA GAA Asp Ala Pro Lys Leu Gly Phe Ser Gln Gly Asp Ile Leu Phe Lys Glu 55 60 65			997
GAG AAG GAA ATT ATC GCC GTT AAT ATC TTG GAT TCT GAA GTC ATT CAC Glu Lys Glu Ile Ile Ala Val Asn Ile Leu Asp Ser Glu Val Ile His 70 75 80			1045
ATC CAA GCT AAG AGC GTG GCA GAA GTA GCG AAA ATA TGC TAT GAA ATA Ile Gln Ala Lys Ser Val Ala Glu Val Ala Lys Ile Cys Tyr Glu Ile 85 90 95			1093
GGA AAC CGC CAT GCG GCT TTA TAC TAT GGC GAG TCT CAA TTT GAA TTT Gly Asn Arg His Ala Ala Leu Tyr Tyr Gly Glu Ser Gln Phe Glu Phe 100 105 110			1141
AAA ACA CCA TTT GAA AAG CCC ACG CTA GCG TTA CTA GAA AAG CTA GGG Lys Thr Pro Phe Glu Lys Pro Thr Leu Ala Leu Leu Glu Lys Leu Gly 115 120 125 130			1189
GTT CAA AAT CGT GTT TTA AGT TCA AAA TTG GAT TCC AAA GAA CGC TTA Val Gln Asn Arg Val Leu Ser Ser Lys Leu Asp Ser Lys Glu Arg Leu 135 140 145			1237
ACC GTG AGC ATG CCC CAT AGT GAG CCT AAT TTT AAG GTC TCA CTG GCG Thr Val Ser Met Pro His Ser Glu Pro Asn Phe Lys Val Ser Leu Ala 150 155 160			1285
AGC GAT TTT AAA GTG GTC ATG AAA TAGAAAAACA ACAA ATG GAT AAA GGA Ser Asp Phe Lys Val Val Met Lys Met Asp Lys Gly 165 170 1			1335
AAA AGC GTG AAA AGC ATT GAA AAA AGC GTG GGT ATG CTC CCA AAA ACT Lys Ser Val Lys Ser Ile Glu Lys Ser Val Gly Met Leu Pro Lys Thr 5 10 15 20			1383
CCA AAG ACA GAC AGC AAT GCT CAT GTG GAT AAT GAA TTT CTG ATT CTG Pro Lys Thr Asp Ser Asn Ala His Val Asp Asn Glu Phe Leu Ile Leu 25 30 35			1431
CAA GTC AAT GAT GCG GTG TTC CCC ATT GGA TCT TAC ACG CAT TCT TTT Gln Val Asn Asp Ala Val Phe Pro Ile Gly Ser Tyr Thr His Ser Phe 40 45 50			1479
GGG CTT TTG GCT AGA AAC TTA CAT CCA GCA AAA AAG GTT ACT AAT AAA Gly Leu Leu Ala Arg Asn Leu His Pro Ala Lys Lys Val Thr Asn Lys			1527

Al  
cont

41  
-7-

55 60 65

GAA AGC GCT TTA AAA TAT TTA AAA GCC AAT CTC TCT AGC CAG TTC CTT 1575  
Glu Ser Ala Leu Lys Tyr Leu Lys Ala Asn Leu Ser Ser Gln Phe Leu  
70 75 80

TAC ACG GAA ATG CTG AGC TTG AAA CTC ACC TAT GAA AGC GCT CTC CAA 1623  
Tyr Thr Glu Met Leu Ser Leu Lys Leu Thr Tyr Glu Ser Ala Leu Gln  
85 90 95 100

CAA GAT TTA AAA AGG ATC TTA GGG GTT GAA GAA ATC ATT ACG CTA TCC 1671  
Gln Asp Leu Lys Arg Ile Leu Gly Val Glu Glu Ile Ile Thr Leu Ser  
105 110 115

ACA AGC CCC ATG GAA TTG CGA TTA GCC AAT CAA AAG CTA GGC AAT CGT 1719  
Thr Ser Pro Met Glu Leu Arg Leu Ala Asn Gln Lys Leu Gly Asn Arg  
120 125 130

TTC ATT AAA ACC TTA CAA GCC ATG AAC GAA TTA GAC ATT GGC GCA TTT 1767  
Phe Ile Lys Thr Leu Gln Ala Met Asn Glu Leu Asp Ile Gly Ala Phe  
135 140 145

TTT AAC GCT TAC GCT CAA CAA ACC GAA GAC CCC ACC CAT GCC ACT AGC 1815  
Phe Asn Ala Tyr Ala Gln Gln Thr Glu Asp Pro Thr His Ala Thr Ser  
150 155 160

TAT GGC GTT TTT GCG GCG AGT TTG GGG ATT GAA TTG AAA AAG GCT TTA 1863  
Tyr Gly Val Phe Ala Ala Ser Leu Gly Ile Glu Leu Lys Lys Ala Leu  
165 170 175 180

AGG CAT TAT CTT TAT GCA CAA ACT TCT AAC ATG GTA ATT AAC TGC GTT 1911  
Arg His Tyr Leu Tyr Ala Gln Thr Ser Asn Met Val Ile Asn Cys Val  
185 190 195

AAA AGC GTC CCA CTA TCT CAA AAC GAT GGG CAA AAA ATC TTA TTG AGC 1959  
Lys Ser Val Pro Leu Ser Gln Asn Asp Gly Gln Lys Ile Leu Leu Ser  
200 205 210

TTG CAA AGC CCT TTT AAC CAG CTC ATA GAA AAA ACC CTA GAA CTA GAC 2007  
Leu Gln Ser Pro Phe Asn Gln Leu Ile Glu Lys Thr Leu Glu Leu Asp  
215 220 225

GAA AGC CAC TTG TGC GCG GCA AGC GTT CAA AAC GAC ATT AAG GCG ATG 2055  
Glu Ser His Leu Cys Ala Ala Ser Val Gln Asn Asp Ile Lys Ala Met  
230 235 240

CAG CAT GAG AGT TTA TAC TCG CGC CTT TAT ATG TCT TGAATTTTAT 2101  
Gln His Glu Ser Leu Tyr Ser Arg Leu Tyr Met Ser  
245 250 255

CTCAAATTGA AAGGAATTTT ATG GTA AAA ATT GGA GTT TGT GGT CCT GTA 2151  
Met Val Lys Ile Gly Val Cys Gly Pro Val

42

42  
-8-

	1	5	10	
	GGA AGC GGT AAA ACC GCC TTG ATT GAA GCT TTA ACG CGC CAC ATG TCA			2199
	Gly Ser Gly Lys Thr Ala Leu Ile Glu Ala Leu Thr Arg His Met Ser			
	15	20	25	
	AAA GAT TAT GAC ATG GCG GTC ATC ACT AAT GAT ATT TAC ACG AAA GAA			2247
	Lys Asp Tyr Asp Met Ala Val Ile Thr Asn Asp Ile Tyr Thr Lys Glu			
	30	35	40	
	GAC GCA GAA TTT ATG TGT AAA AAT TCG GTG ATG CCA CGA GAG AGG ATC			2295
	Asp Ala Glu Phe Met Cys Lys Asn Ser Val Met Pro Arg Glu Arg Ile			
	45	50	55	
	ATT GGC GTA GAA ACA GGA GGC TGT CCG CAC ACG GCT ATT AGA GAA GAC			2343
	Ile Gly Val Glu Thr Gly Gly Cys Pro His Thr Ala Ile Arg Glu Asp			
	60	65	70	
	GCT TCT ATG AAT TTA GAA GCC GTA GAA GAA ATG CAT GGC CGT TTC CCT			2391
	Ala Ser Met Asn Leu Glu Ala Val Glu Glu Met His Gly Arg Phe Pro			
	75	80	85	90
	AAT TTG GAA TTG CTT TTG ATT GAA AGC GGA GGC AGT AAC CTT TCA GCG			2439
	Asn Leu Glu Leu Leu Leu Ile Glu Ser Gly Gly Ser Asn Leu Ser Ala			
	95	100	105	
	ACT TTC AAC CCA GAG CTA GCG GAC TTT ACG ATC TTT GTG ATT GAT GTG			2487
	Thr Phe Asn Pro Glu Leu Ala Asp Phe Thr Ile Phe Val Ile Asp Val			
	110	115	120	
	GCT GAG GGC GAT AAA ATC CCC AGA AAA GGC GGG CCA GGA ATC ACG CGT			2535
	Ala Glu Gly Asp Lys Ile Pro Arg Lys Gly Gly Pro Gly Ile Thr Arg			
	125	130	135	
	TCA GAC TTG CTT GTC ATC AAT AAG ATT GAT TTA GCC CCC TAT GTG GGA			2583
	Ser Asp Leu Leu Val Ile Asn Lys Ile Asp Leu Ala Pro Tyr Val Gly			
	140	145	150	
	GCC GAC TTG AAA GTC ATG GAA AGG GAT TCT AAA AAA ATC GCG GCG AAA			2631
	Ala Asp Leu Lys Val Met Glu Arg Asp Ser Lys Lys Ile Ala Ala Lys			
	155	160	165	170
	AGC CCT TTA TTT TTA CCG AAT ATC CGC GCT AAA GAA GGT TTA GAC GAT			2679
	Ser Pro Leu Phe Leu Pro Asn Ile Arg Ala Lys Glu Gly Leu Asp Asp			
	175	180	185	
	GTG ATC GCT TGG ATC AAG CGC AAC GCT TTA TTG GAA GAT TG ATG AAC			2726
	Val Ile Ala Trp Ile Lys Arg Asn Ala Leu Leu Glu Asp Met Asn			
	190	195	1	
	ACT TAC GCT CAA GAA TCC AAG CTC AGG TTA AAA ACC AAA ATA GGG GCT			2774
	Thr Tyr Ala Gln Glu Ser Lys Leu Arg Leu Lys Thr Lys Ile Gly Ala			

all  
cont

43

43  
-9-

	5	10	15	
	GAC GGG CGG TGC GTG ATT GAA GAC AAT TTT TTC ACG CCC CCC TTT AAG			2822
	Asp Gly Arg Cys Val Ile Glu Asp Asn Phe Phe Thr Pro Pro Phe Lys			
	20	25	30	
	CTC ATG GCG CCC TTT TAC CCT AAA GAC GAT TTA GCG GAA ATC ATG CTT			2870
	Leu Met Ala Pro Phe Tyr Pro Lys Asp Asp Leu Ala Glu Ile Met Leu			
	35	40	45	50
	TTA GCG GTA AGC CCT GGC TTA ATG AAA GGC GAT GCA CAA GAT GTG CAA			2918
	Leu Ala Val Ser Pro Gly Leu Met Lys Gly Asp Ala Gln Asp Val Gln			
		55	60	65
	TTG AAC ATC GGT CCA AAT TGC AAG TTA AGG ATC ACT TCG CAA TCC TTT			2966
	Leu Asn Ile Gly Pro Asn Cys Lys Leu Arg Ile Thr Ser Gln Ser Phe			
		70	75	80
	GAA AAA ATC CAT AAC ACT GAA GAC GGG TTT GCT AGC AGA GAC ATG CAT			3014
	Glu Lys Ile His Asn Thr Glu Asp Gly Phe Ala Ser Arg Asp Met His			
		85	90	95
	ATC GTT GTG GGG GAA AAC GCT TTT TTA GAC TTC GCG CCC TTC CCG TTA			3062
	Ile Val Val Gly Glu Asn Ala Phe Leu Asp Phe Ala Pro Phe Pro Leu			
	100	105	110	
	ATC CCC TTT GAA AAC GCG CAT TTT AAG GGC AAT ACC ACG ATT TCT TTG			3110
	Ile Pro Phe Glu Asn Ala His Phe Lys Gly Asn Thr Thr Ile Ser Leu			
	115	120	125	130
	CGC TCT AGC TCC CAA TTG CTC TAT AGT GAA ATC ATT GTC GCA GGG CGA			3158
	Arg Ser Ser Ser Gln Leu Leu Tyr Ser Glu Ile Ile Val Ala Gly Arg			
		135	140	145
	GTG GCG CGC AAT GAG TTG TTT AAA TTC AAC CGC TTG CAC ACC AAA ATC			3206
	Val Ala Arg Asn Glu Leu Phe Lys Phe Asn Arg Leu His Thr Lys Ile			
	150	155	160	
	TCT ATT TTA CAA GAT GAG AAA CCC ATC TAT TAT GAC AAC ACG ATT TTA			3254
	Ser Ile Leu Gln Asp Glu Lys Pro Ile Tyr Tyr Asp Asn Thr Ile Leu			
	165	170	175	
	GAT CCC AAA ACC ACC GAC TTA AAT AAC ATG TGC ATG TTT GAT GGC TAT			3302
	Asp Pro Lys Thr Thr Asp Leu Asn Asn Met Cys Met Phe Asp Gly Tyr			
	180	185	190	
	ACG CAT TAT TTG AAT TTG GTG CTG GTC AAT TGC CCC ATA GAG CTG TCT			3350
	Thr His Tyr Leu Asn Leu Val Leu Val Asn Cys Pro Ile Glu Leu Ser			
	195	200	205	210
	GGC GTG CGA GGA TTG ATT GAA GAG AGC GAA GGA GTG GAT GGA GCC GTG			3398
	Gly Val Arg Gly Leu Ile Glu Glu Ser Glu Gly Val Asp Gly Ala Val			

44

215	220	225	
AGT GAA ATC GCT AGT TCT CAT TTA TGC CTG AAA GCT TTA GCG AAA GGC			3446
Ser Glu Ile Ala Ser Ser His Leu Cys Leu Lys Ala Leu Ala Lys Gly			
230	235	240	
TCA GAA CCC TTG TTG CAT TTA AGA GAA AAA ATC GCT CGC TTT ATC ACG			3494
Ser Glu Pro Leu Leu His Leu Arg Glu Lys Ile Ala Arg Phe Ile Thr			
245	250	255	
CAA ACG ATT ACG CCA AAG GTT TAAAAAACAC TTAAAAAAG ATTATACCCT			3545
Gln Thr Ile Thr Pro Lys Val			
260	265		
TTAGTCTTTT TTAA			3559

(2) INFORMATION FOR SEQ ID NO:2:

*at cont*

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 5 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Leu Phe Ser Ile Phe  
 1 5

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 195 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Met Leu Gly Leu Val Leu Leu Tyr Val Gly Ile Val Leu Ile Ser Asn  
 1 5 10 15

Gly Ile Cys Gly Leu Thr Lys Val Asp Pro Lys Ser Thr Ala Val Met  
 20 25 30

Asn Phe Phe Val Gly Gly Leu Ser Ile Ile Cys Asn Val Val Ile  
 35 40 45



45  
-11-

Thr Tyr Ser Ala Leu Asn Pro Thr Ala Pro Val Glu Gly Ala Glu Asp  
50 55 60

Ile Ala Gln Val Ser His His Leu Thr Asn Phe Tyr Gly Pro Ala Thr  
65 70 75 80

Gly Leu Leu Phe Gly Phe Thr Tyr Leu Tyr Ala Ala Ile Asn His Thr  
85 90 95

Phe Gly Leu Asp Trp Arg Pro Tyr Ser Trp Tyr Ser Leu Phe Val Ala  
100 105 110

Ile Asn Thr Ile Pro Ala Ala Ile Leu Ser His Tyr Ser Asp Met Leu  
115 120 125

Asp Asp His Lys Val Leu Gly Ile Thr Glu Gly Asp Trp Trp Ala Ile  
130 135 140

Ile Trp Leu Ala Trp Gly Val Leu Trp Leu Thr Ala Phe Ile Glu Asn  
145 150 155 160

Ile Leu Lys Ile Pro Leu Gly Lys Phe Thr Pro Trp Leu Ala Ile Ile  
165 170 175

Glu Gly Ile Leu Thr Ala Trp Ile Pro Ala Trp Leu Leu Phe Ile Gln  
180 185 190

His Trp Val  
195

Al  
cont

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 170 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Ile Ile Glu Arg Leu Ile Gly Asn Leu Arg Asp Leu Asn Pro Leu  
1 5 10 15

Asp Phe Ser Val Asp Tyr Val Asp Leu Glu Trp Phe Glu Thr Arg Lys  
20 25 30

Lys Ile Ala Arg Phe Lys Thr Arg Gln Gly Lys Asp Ile Ala Val Arg  
35 40 45

Leu Lys Asp Ala Pro Lys Leu Gly Phe Ser Gln Gly Asp Ile Leu Phe

46

46  
-12-

50 55 60

Lys Glu Glu Lys Glu Ile Ile Ala Val Asn Ile Leu Asp Ser Glu Val  
65 70 75 80

Ile His Ile Gln Ala Lys Ser Val Ala Glu Val Ala Lys Ile Cys Tyr  
85 90 95

Glu Ile Gly Asn Arg His Ala Ala Leu Tyr Tyr Gly Glu Ser Gln Phe  
100 105 110

Glu Phe Lys Thr Pro Phe Glu Lys Pro Thr Leu Ala Leu Leu Glu Lys  
115 120 125

Leu Gly Val Gln Asn Arg Val Leu Ser Ser Lys Leu Asp Ser Lys Glu  
130 135 140

Arg Leu Thr Val Ser Met Pro His Ser Glu Pro Asn Phe Lys Val Ser  
145 150 155 160

Leu Ala Ser Asp Phe Lys Val Val Met Lys  
165 170

Al  
cont

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 256 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Met Asp Lys Gly Lys Ser Val Lys Ser Ile Glu Lys Ser Val Gly Met  
1 5 10 15

Leu Pro Lys Thr Pro Lys Thr Asp Ser Asn Ala His Val Asp Asn Glu  
20 25 30

Phe Leu Ile Leu Gln Val Asn Asp Ala Val Phe Pro Ile Gly Ser Tyr  
35 40 45

Thr His Ser Phe Gly Leu Leu Ala Arg Asn Leu His Pro Ala Lys Lys  
50 55 60

Val Thr Asn Lys Glu Ser Ala Leu Lys Tyr Leu Lys Ala Asn Leu Ser  
65 70 75 80

Ser Gln Phe Leu Tyr Thr Glu Met Leu Ser Leu Lys Leu Thr Tyr Glu  
85 90 95

47

47  
-18-

Ser Ala Leu Gln Gln Asp Leu Lys Arg Ile Leu Gly Val Glu Glu Ile  
 100 105 110  
 Ile Thr Leu Ser Thr Ser Pro Met Glu Leu Arg Leu Ala Asn Gln Lys  
 115 120 125  
 Leu Gly Asn Arg Phe Ile Lys Thr Leu Gln Ala Met Asn Glu Leu Asp  
 130 135 140  
 Ile Gly Ala Phe Phe Asn Ala Tyr Ala Gln Gln Thr Glu Asp Pro Thr  
 145 150 155 160  
 His Ala Thr Ser Tyr Gly Val Phe Ala Ala Ser Leu Gly Ile Glu Leu  
 165 170 175  
 Lys Lys Ala Leu Arg His Tyr Leu Tyr Ala Gln Thr Ser Asn Met Val  
 180 185 190  
 Ile Asn Cys Val Lys Ser Val Pro Leu Ser Gln Asn Asp Gly Gln Lys  
 195 200 205  
 Ile Leu Leu Ser Leu Gln Ser Pro Phe Asn Gln Leu Ile Glu Lys Thr  
 210 215 220  
 Leu Glu Leu Asp Glu Ser His Leu Cys Ala Ala Ser Val Gln Asn Asp  
 225 230 235 240  
 Ile Lys Ala Met Gln His Glu Ser Leu Tyr Ser Arg Leu Tyr Met Ser  
 245 250 255

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 199 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met Val Lys Ile Gly Val Cys Gly Pro Val Gly Ser Gly Lys Thr Ala  
 1 5 10 15  
 Leu Ile Glu Ala Leu Thr Arg His Met Ser Lys Asp Tyr Asp Met Ala  
 20 25 30  
 Val Ile Thr Asn Asp Ile Tyr Thr Lys Glu Asp Ala Glu Phe Met Cys  
 35 40 45

48

48  
-14-

Lys Asn Ser Val Met Pro Arg Glu Arg Ile Ile Gly Val Glu Thr Gly  
50 55 60

Gly Cys Pro His Thr Ala Ile Arg Glu Asp Ala Ser Met Asn Leu Glu  
65 70 75 80

Ala Val Glu Glu Met His Gly Arg Phe Pro Asn Leu Glu Leu Leu Leu  
85 90 95

Ile Glu Ser Gly Gly Ser Asn Leu Ser Ala Thr Phe Asn Pro Glu Leu  
100 105 110

Ala Asp Phe Thr Ile Phe Val Ile Asp Val Ala Glu Gly Asp Lys Ile  
115 120 125

Pro Arg Lys Gly Gly Pro Gly Ile Thr Arg Ser Asp Leu Leu Val Ile  
130 135 140

Asn Lys Ile Asp Leu Ala Pro Tyr Val Gly Ala Asp Leu Lys Val Met  
145 150 155 160

Glu Arg Asp Ser Lys Lys Ile Ala Ala Lys Ser Pro Leu Phe Leu Pro  
165 170 175

Asn Ile Arg Ala Lys Glu Gly Leu Asp Asp Val Ile Ala Trp Ile Lys  
180 185 190

Arg Asn Ala Leu Leu Glu Asp  
195

A!  
cont

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 265 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Met Asn Thr Tyr Ala Gln Glu Ser Lys Leu Arg Leu Lys Thr Lys Ile  
1 5 10 15

Gly Ala Asp Gly Arg Cys Val Ile Glu Asp Asn Phe Phe Thr Pro Pro  
20 25 30

Phe Lys Leu Met Ala Pro Phe Tyr Pro Lys Asp Asp Leu Ala Glu Ile  
35 40 45

Met Leu Leu Ala Val Ser Pro Gly Leu Met Lys Gly Asp Ala Gln Asp

49

49  
-15-

50		55		60											
Val 65	Gln	Leu	Asn	Ile	Gly 70	Pro	Asn	Cys	Lys	Leu 75	Arg	Ile	Thr	Ser	Gln 80
Ser	Phe	Glu	Lys	Ile 85	His	Asn	Thr	Glu	Asp 90	Gly	Phe	Ala	Ser	Arg 95	Asp
Met	His	Ile	Val 100	Val	Gly	Glu	Asn	Ala 105	Phe	Leu	Asp	Phe	Ala 110	Pro	Phe
Pro	Leu	Ile 115	Pro	Phe	Glu	Asn	Ala 120	His	Phe	Lys	Gly	Asn 125	Thr	Thr	Ile
Ser	Leu	Arg	Ser	Ser	Ser	Gln	Leu	Leu	Tyr	Ser	Glu 140	Ile	Ile	Val	Ala
Gly 145	Arg	Val	Ala	Arg	Asn 150	Glu	Leu	Phe	Lys	Phe 155	Asn	Arg	Leu	His	Thr 160
Lys	Ile	Ser	Ile 165	Leu	Gln	Asp	Glu	Lys	Pro 170	Ile	Tyr	Tyr	Asp	Asn 175	Thr
Ile	Leu	Asp	Pro 180	Lys	Thr	Thr	Asp	Leu	Asn 185	Asn	Met	Cys	Met 190	Phe	Asp
Gly	Tyr	Thr	His	Tyr	Leu	Asn	Leu	Val	Leu	Val	Asn 205	Cys	Pro	Ile	Glu
Leu 210	Ser	Gly	Val	Arg	Gly	Leu	Ile	Glu	Glu	Ser	Glu 220	Gly	Val	Asp	Gly
Ala 225	Val	Ser	Glu	Ile	Ala 230	Ser	Ser	His	Leu	Cys 235	Leu	Lys	Ala	Leu	Ala 240
Lys	Gly	Ser	Glu	Pro	Leu	Leu	His	Leu	Arg 250	Glu	Lys	Ile	Ala	Arg 255	Phe
Ile	Thr	Gln	Thr	Ile	Thr	Pro	Lys	Val 265							

1  
a  
cont

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 33 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: unknown
  - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GCGAAAATAT GCTATGAAAT AGGAAACCGC CAT

33

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 11 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

*Ala  
cont*

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Ala	Lys	Ile	Cys	Tyr	Glu	Ile	Gly	Asn	Arg	His
1				5					10	

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 16 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: unknown
  - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 16
- (D) OTHER INFORMATION: /note= "W is either A or T."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

TGGYAYRNNN NYYG CW

16

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 6 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: unknown
  - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

51  
-11-

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

TTGACA

6

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 6 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: unknown
  - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

TATAAT

6

52